citedkb

WHITE PAPER

Keeping company knowledge correct, not just searchable

How CitedKB finds and fixes problems in an organization's knowledge, answers from it with sources, and keeps every document private to the people allowed to read it.

16 November 2026 citedkb.com contact@citedkb.com

Summary

Most companies now have an AI tool that can answer questions from their documents. Very few have anything that checks whether those documents are **right**.

When documents contradict each other, are out of date or leave questions unanswered, an AI assistant repeats those problems quickly and confidently, and to everyone at once. The bigger the document collection grows, the worse this gets.

CitedKB is built around a different idea: the quality of answers depends on the quality of the knowledge behind them. It does two jobs together:

- **It keeps company knowledge healthy.** It finds contradictions, outdated documents and gaps, and helps the people who own the information fix them. AI suggests; people decide.
- **It makes that knowledge easy to use.** Everyone can ask questions and get answers that show their sources. Experts verify the most important answers, and the same verified answer is then given to everyone.

The same care goes beyond documents: **data** is calculated from every row and watched for anomalies, and **scanned sites and sensors** are compared over time to show what changed. Statistics find the problems; the AI explains them.

All of this follows one strict rule: **nobody ever sees, or gets answers from, a document they are not allowed to read**, including the results of the quality checks themselves.

---

1. The problem

Company knowledge decays

Policies are updated, but old copies stay in shared folders. Two departments write their own version of the same rule. A draft is saved next to the approved version. A document that was correct last year quietly becomes wrong, and nobody notices until someone acts on it.

AI makes this worse, not better

An AI assistant answers from whatever text it finds. If it finds the old version, it gives the old answer, confidently. If it finds two versions that disagree, it may pick one at random, or mix them. Because the answer sounds right, people trust it.

It gets worse as you grow

With a few hundred documents, a search usually finds the right text. With tens of thousands, it often finds text that is **similar but wrong**: an old version, a related but different policy, or a duplicate pushing out the useful passage. This is the most common reason AI tools give wrong answers in large companies. The AI isn't inventing things from nothing; it's being handed the wrong text.

Access control is usually an afterthought

Departments hold sensitive information. Many tools check permissions when they search, but leak information elsewhere: suggested answers, review queues, notifications, audit logs, or the AI's own notes from earlier answers.

---

2. How CitedKB compares

The tools companies use today

Kind of tool	What it does well	What's usually missing
AI assistants built into office suites, and general enterprise chat assistants	Fast, natural answers across email, files and chats; help with writing and everyday tasks	They answer from whatever they find. Checking whether company knowledge is correct, and getting the owners to fix it, is not what they are built for
Enterprise search	Finds documents by keyword or meaning, across many systems	Doesn't judge whether a document is still correct, or whether two documents disagree
Wikis and intranets	A curated place for "the official version"	Depends on people keeping it up to date by hand, and quickly falls behind
Document management systems	Versions, approvals, retention	Can't tell when two different documents contradict each other

The gap: none of these, on its own, connects **finding** problems in the knowledge, **fixing** them with the right owners, and **answering** from the fixed knowledge, all under one permission model that also covers the quality checks themselves.

CitedKB compared with office-suite AI assistants

AI assistants built into office suites are excellent personal productivity tools. CitedKB is built for a different job: **making the organization's knowledge trustworthy**, and keeping it that way. The two work well together: CitedKB connects to AI assistants through MCP (Model Context Protocol), so people can use CitedKB's checked knowledge from inside the assistant they already use.

	CitedKB	Typical office-suite AI assistant
Main purpose	Keep company knowledge correct, then answer from it	Help each person work faster across their files, email and meetings
Contradictions between documents	Scans compare documents and show conflicting passages side by side	Not a built-in part of the product; it answers from what it finds
Outdated documents and duplicates	Found automatically; duplicates don't crowd out answers	Usually not checked
Fixing the source	Contradictions go to the departments that own the information, with deadlines, escalation and a recorded decision	No built-in process for agreeing on and recording the correct version

	CitedKB	Typical office-suite AI assistant
Rejected information coming back	Caught when a rejected statement is uploaded again	Not tracked
Same question, same answer	Expert-verified answers, given to everyone who asks, re-checked when sources change	Answers are generated each time and may vary
Knowledge gaps	Questions no document can answer, grouped by topic	Usually not reported
Permissions for quality checks	Checks follow sharing; "read only" keeps a document out of others' checks; private documents get a private check	Not applicable (no quality checks)
Private documents	Never leak through answers, findings, review queues, AI notes or the audit log	Follows the permissions of the underlying file storage
Numbers from data	Database queries over every row; personal-data columns hidden from the AI	Varies by product and file type
Point clouds, sensors and anomalies	Scan monitoring and statistical anomaly detection built in	Not part of a general assistant
Research-grade reports	Systematic reviews and research papers with academic citation styles and reference files	General writing help
Where data lives	Cloud infrastructure in Europe, with an optional dedicated cloud account per organization	Depends on the vendor's cloud
Cost control	Limits per organization, department and person; work pauses and resumes instead of failing; past comparisons reused	Usually a fixed price per user
Tied to one office suite?	No: uploads, cloud drives, websites, databases and your own systems	Usually tied to its own suite

Comparisons describe the typical capabilities of each kind of tool in late 2026, not any one product. Products change often.

What sets CitedKB apart

- **It checks the knowledge, not just the question.** Contradiction, outdated-document and duplicate detection is the core of the product, not an add-on.
- **It closes the loop with people.** Findings go to the owners, decisions are recorded, and AI only suggests.
- **Verified answers build up over time.** Each expert check makes the system more consistent for everyone.
- **Permissions cover everything, including the checks.** Checks follow sharing; reading and scanning are separate permissions; private stays private, down to the audit log.
- **It stays accurate as you grow.** Reranking, duplicate handling and health-aware search fight the drop in quality that large collections cause.
- **It's cost-aware by design.** Document pairs are compared once and the result is shared by every check that needs it.

- **Statistics decide, AI explains.** In data and sensor monitoring, problems are found by measurable statistics.
- **It goes beyond documents.** Data, databases, point-cloud scans and sensors live in the same governed system.
- **European hosting**, with a dedicated account option for organizations that need strong separation.
- **It works with what you already use**, including chat apps, AI assistants, cloud drives and your own systems, without replacing them.

---

3. What CitedKB does

Knowledge health: finding what's wrong

A scan reads your documents carefully and compares them with each other. It finds:

- **Contradictions:** two documents say different things about the same subject, shown side by side, with which one is currently in effect.
- **Outdated documents:** replaced by newer rules, or not reviewed for a long time.
- **Duplicates:** copies of the same text that confuse readers and AI alike.
- **Knowledge gaps:** questions people ask that no document can answer, grouped by topic.

Every new upload is checked automatically against what the uploader's team already has.

Fixing it, with the right people

- Each contradiction goes to the departments that own the information, with a suggested decision-maker, deadlines and escalation.
- AI suggests a fix, such as which version to keep or a corrected copy. **A person always decides.**
- Decisions are remembered: if a rejected statement comes back in a new upload, it's caught.
- Review schedules show which documents are due for checking.

Answers you can check

- Every answer shows its sources: the document, the page and the picture it came from.
- Answers use the newest version of each document by default.
- The AI is told about known contradictions and which version was agreed to be correct.
- Hidden and corrected documents are respected once a fix is made.
- Questions can be asked in any language, even when the documents are in another language.

Every page read properly

Answers are only as good as the text behind them. CitedKB reads every document **page by page, the way a person would**: it looks at each page together with its own text, so tables keep their rows and columns, scanned pages and photos of forms are read, and charts and diagrams are

described. Word files are read directly from the file, by heading, with their pictures described. Empty pages are skipped.

- **Document types:** each department has a type (General, Finance, Legal and policies, Technical, Forms and scans, or the organization's own), and any document can have its own. The type decides how pages are read and how the text is split for search: a contract keeps its clauses together, a form is split into small pieces.
- **High accuracy** reads a difficult document again, with more care, when someone asks.
- **Never read twice:** when a document changes, only the pages that changed are read again. Every page is read once and reused, so nobody pays twice for the same page.
- Citations name the page, the slide or the section of a Word document.

Verified answers

For the questions that matter most, an expert checks the answer once. After that, everyone who asks the same question, in any wording, gets the same verified answer, clearly marked. When a source changes or is deleted, the answer goes back for review automatically.

Numbers that are counted, not guessed

Spreadsheets and databases become datasets. You ask in plain language, the AI writes a database query, runs it on **every row**, and answers with a table or chart you can check. Columns with personal data (such as national identity numbers, bank account numbers and card numbers) are detected and hidden, and never reach the AI.

Answers are short and to the point, and look at the whole result: a fifty-year trend is described from its first year to its last, not from the first few rows. Charts include long time series as lines (up to 2,000 points), dates as years or months, axis titles in plain words, and rates shown as percentages.

Anomaly detection in your data

The AI picks what is worth watching in a dataset. **Statistics and machine learning** then check every row for unusual periods, sudden changes in level, groups that stand out, and extreme or unusual combinations of values. Checks run automatically or on a schedule, and serious findings can be emailed (only to people allowed to see the data) or sent to your own systems. **The AI chooses what to watch; the maths decides what is unusual.**

Python analysis in a safe sandbox

Python code runs on your datasets in an isolated sandbox with no internet, no other data, and personal-data columns left out. Paste your own code, or ask the AI to write and change it. Results come back as tables, charts and files.

Scan monitoring: point clouds and sensors

- **3D scans** (LiDAR point clouds) and **2D scans** (profile and line scanners) are kept as a history per location, such as a tunnel section, bridge or station.
- Each new scan is compared with the **baseline** and with the **scan before it**: what appeared, what's gone, where and by how much, in a 3D viewer with changes shown in colour.

- **Sensor data** is checked for spikes, level shifts, real trends (statistically tested, not noise), drift against the baseline, and sensors that went missing or appeared.
- **Reports** compare any versions you choose, as a PDF with findings, heat maps and a trend page.
- The AI writes the report summary **only from the measured facts**; it never sees the raw points. If scans didn't line up well, the report says so.

Reports and announcements

CitedKB writes research papers, systematic and literature reviews, topic reports, briefings, comparisons and summaries from the documents you choose, with a citation for every claim in standard academic styles. Admins and managers can also send announcements, written by the AI only from their notes, as a progress update, newsletter, FAQ or general update.

Every file is made ready to share: Word and PDF files have page numbers and tables that repeat their header row on each page, Excel files print with their header on every page, and PowerPoint files come in one of five designs, or in your own colours, light or dark, and font.

Use it where people already work

The same knowledge base can be used from chat apps, from AI assistants, from cloud drives and websites that stay in sync, and from your own systems through an API. Every channel follows exactly the same permission rules.

Integrations the organization controls

- **Technical setup and document access are separate.** Admins and an integration team connect company accounts and create API keys; they never see documents. Department managers choose which folders a department syncs.
- **Every API key belongs to one department, not a person.** It keeps working when people leave. It can send documents and change or delete only the documents it sent itself. It can only search or ask questions after the department's manager approves.
- **Sharing stays inside the department.** Each document sent through the API can be shared with the department, a team or named people in it. People who are not in CitedKB are never given access: the answer lists who was added and who was not.
- **Integration tokens are the department's.** Reading documents from cloud drives and the API uses each department's own integration tokens, set by the admins and shown on their own line in usage. No employee pays for an import. When tokens run out, documents wait and continue when tokens are back; none are lost.
- Documents from cloud drives and the API are read just as carefully as uploads, with their document type.

4. What the AI can produce

You ask for...	What you get
An answer to a question	A clear answer with citations to document, page and picture
A summary, analysis, detailed write-up, steps, checklist, comparison or email	Exactly that form, written from your documents
A table	A table built from the figures in your documents
A chart from documents	The figures as a table ready to chart, with what it shows; missing figures are never invented
Counts, totals, rankings and charts from data	Calculated from every row, with a table or chart you can check
Custom calculations	Python results: tables, charts and files
A full report	Research paper, review, topic report, briefing, comparison, summary or your own structure, with citations, as Word, PowerPoint, PDF or Excel, with charts from your datasets and page numbers
A chat or answer as a file	PDF, Word or PowerPoint with the sources; PowerPoint in the slide design you choose
A report from a conversation	The questions, answers and sources of a chat, written up with one click
What's unusual in data	A findings report on unusual periods, shifts and values
What changed at a scanned location	A PDF with an AI summary, findings, heat maps and trends
An announcement	A progress update, newsletter, FAQ or general update from your notes
An answer about a picture or a file	The AI reads the attached picture or file and searches your documents for it

The AI also understands what's inside your files: diagrams as shapes and flows, presentations including the numbers behind charts and the speaker notes, and pictures inside documents.

5. Built-in trust

Every document has an owner

The person who uploads a document owns it and chooses who can see it: only themselves (private), specific people, teams, departments, or the whole organization. **Viewers** read it, **editors** can also upload new versions, and **owners** can do everything.

Reading and scanning are separate

A document can be shared **"read only"**: people can read it, but it's left out of their knowledge-health checks.

Knowledge health follows sharing

Sharing	Included in
Private, or shared with chosen people	Only the owner's own "My documents" check
A team	That team's check
A department	That department's check, and the checks of its teams
The whole organization	Every check

A finding is shown only to people who can read every document in it. When sharing is narrowed, the finding disappears for people who lost access immediately.

Asking for access

To read another department's documents, people ask for access. The department's manager decides, a team lead decides for their team's documents, and the documents' owner decides when they own everything asked for.

Private stays private, everywhere

A private document's content and name never leak through answers, search results, verified answers, findings, notifications, expert review queues, the AI's notes from earlier answers, or the audit log. Only the owner sees its name there.

Security and hosting

- Hosted in Europe. The AI runs inside the same cloud, which does not train models on your data. One step, ranking the passages found for a question, runs in Frankfurt; it also stays in Europe.
- Customer data is kept apart from the service itself, and organizations can choose a dedicated cloud account with its own encryption key.
- Permission decisions are always made on the server, never in the browser.
- Single sign-on, two-step verification, and encryption in transit and at rest.
- Databases can be restored to any minute in the last 35 days.
- GDPR tools: export or erase a person's data, with clear retention rules.

More details are on our **security page**.

---

6. Built to stay accurate as you grow

- **Every page read the way its type needs:** tables, scans, pictures and long clauses keep their meaning in search.
- **Reranking:** more candidate passages are fetched, then a ranking model orders them by how well they answer the question. Only the best are used.
- **One copy of duplicates:** known copies take one place, so other relevant information gets in.

- **Settled problems are respected:** hidden documents are left out, corrected copies used, and agreed decisions passed to the AI.
- **Latest version by default**, so old versions don't crowd out new ones.
- **Strict answering rules:** the AI answers only from its sources, and says "not found" rather than guessing.
- **Efficient checks:** pairs of documents are compared once per version, and every later check reuses the result without asking the AI again. A document deleted during a scan is left out at once.

7. What CitedKB fixes

Problem	How CitedKB fixes it
Conflicting policies across departments	Finds the contradiction, sends it to the owners, records the decision
AI answers based on old documents	Latest-version answers, outdated documents found, corrected copies used
"Which version is right?"	Shows both passages, which is in effect, and who decided
The same question answered differently	Expert-verified answers, the same for everyone
Nobody knows what's missing	Knowledge gaps grouped by topic, from real questions
Wrong numbers from spreadsheets	Real queries on every row, with tables and charts you can check
Sensitive documents leaking through AI	One permission model across answers, checks, reviews, notes and logs
Answer quality falling as content grows	Reranking, duplicate handling, health-aware search
Unpredictable AI costs	Limits per organization, department and person; comparisons reused; pages never read twice
Scanned PDFs, tables and diagrams answered badly	Every page read like a person would, with the document's type
Imports eating into people's tokens	Integration tokens per department, on their own line in usage
Problems in data found too late	Automatic and scheduled anomaly checks with alerts
Changes at sites missed between inspections	Every scan compared with the baseline and the previous scan
Sensor drift and failures going unnoticed	Spikes, shifts, trends, drift and missing sensors detected statistically

8. Who it's for

- Organizations with many departments and information that must stay inside them: public bodies, healthcare, finance and legal teams.

- Teams whose documents change often: policies, procedures, technical documentation, compliance.
- Infrastructure and asset operators who scan tunnels, bridges, stations or equipment.
- Teams that work with data and need trustworthy numbers and early warnings.
- Anyone who has tried an AI assistant and found it confidently wrong.

---

9. What we're working on next

These are directions we're actively developing: plans, not promises of exact features or dates.

- **Answer checking before an answer is shown:** each sentence compared with its sources; unsupported sentences removed or marked.
- **A quality bar for search results:** when no passage truly answers a question, CitedKB says so instead of answering from weak matches.
- **Continuous quality measurement:** a standing set of real questions, re-run regularly, so any drop is spotted early.
- **Smarter use of document age:** older documents rank slightly lower on the same subject, without being hidden.
- **"My documents" across departments:** private documents checked against everything a person can read.
- **Clearer ownership of shared knowledge:** contradictions routed to the right decision-maker automatically.
- **Charts directly in answers**, next to the tables answers give today.
- **Comparing images in scan monitoring**, next to point clouds and sensor data.

---

10. Conclusion

AI makes company knowledge easy to reach. It doesn't make it correct. CitedKB treats knowledge as something that needs care: **check it, fix it with the people who own it, then answer from it**, with sources, verified answers and permissions that hold everywhere.

Correct knowledge first. Fast answers second. Privacy always.

Want to see it with your own documents? **Contact us**.